

# Henry Xiao

San Francisco, CA | [henryxiao2001@gmail.com](mailto:henryxiao2001@gmail.com) | [linkedin.com/in/henry-xiao-hx](https://www.linkedin.com/in/henry-xiao-hx) | [github.com/Henry-Xiao-HX](https://github.com/Henry-Xiao-HX)

---

## SUMMARY:

Data and AI Engineering leader specializing in Agentic AI, enterprise data modernization, and responsible AI adoption. Founded and scaled IBM's Financial Services Market Agentic Center of Excellence, driving reusable agentic solutions across 15+ products and enabling adoption among 300+ practitioners. Proven track record of translating emerging AI capabilities into scalable solutions that deliver measurable business impact.

---

## PROFESSIONAL EXPERIENCE:

### Data & AI Customer Success Engineer, Agentic CoE Tech Lead, IBM

Jan 2025 - present

- Founded and scaled the IBM FSM CSE Agentic Center of Excellence, coordinating 10+ contributors to build 25+ reusable agentic AI patterns across 15+ IBM products, transforming an idea into an organization-wide initiative that accelerates AI enablement across the IBM Financial Services Market community and reaching 300+ technical sellers, engineers, and executives.
- Developed a multi-agent AI system leveraging the watsonx MCP server to optimize pipeline design, data quality management, and root cause analysis while dynamically balancing execution speed, cloud cost, and resource utilization. The work is featured by IBM leadership in the quarterly technical recognition announcement.
- Designed a context-aware knowledge management framework inspired by Karpathy's LLM Wiki pattern, combining organizational knowledge and persistent agent memory to streamline daily execution; presented at IBM's AI Think Tank and subsequently adopted by a worldwide audience.
- Led collaboration with client team's architects and engineers to map end-to-end legacy data pipeline, implementing LLM-enabled metadata enrichment (with Granite models), and automated data quality remediation workflows. These efforts delivered a gen-ai-ready data foundation with a projected \$6M annual savings (\$1.2M direct labor savings, \$1M on data error reduction, and \$2.4M on data quality improvements).

### AI Engineer - IBM

May 2024 - Jan 2025

- Fine-tuned Llama model leveraging InstrutLab and Large-scale Alignment for Chatbot (LAB) method, generating synthetic data with a teacher model, and performing Parameter-Efficient-Fine-Tuning (PEFT) to inject IBM Risk Atlas domain expertise into the open-source generative AI model.
- Built a Retrieval-Augmented-Generation (RAG) pipeline using Elasticsearch as the vector store with Sentence Transformers as the embedding model for unstructured Slack data; Implemented entity-extraction and taxonomy-based categorization using Llama to additionally support Text-to-SQL. The streamlined knowledge retrieval yielded an estimated 8 hours/week productivity gains.

### Associate AI Engineer - IBM

Aug 2023 - May 2024

- Designed machine learning model evaluation pipelines, assessing fairness, data quality (accuracy, precision), output drift, and feature drift, integrating into watsonx to ensure enterprise AI compliance.
- 

## EDUCATION & CERTIFICATION

Georgia Institute of Technology, M.S. Computer Science

In Progress

University of California, San Diego

Sep 2020 - June 2023

- B.S. Mathematics-Computer Science, GPA 3.86

IBM AI Product Manager Professional Certificate

Feb 2026

---

## Skills:

- Agentic AI & Generative AI: LLM, AI Agents, RAG, Context Engineering, ReAct, MCP, PEFT, AI Governance
- Enterprise Data Platforms: IBM watsonx, Data Integration Architectures, Data Governance
- Engineering: Python, Java, Git, API Design, Solution Architecture, Business Value Articulation